

# Spatial Pattern Analysis of Hop Powdery Mildew in the Pacific Northwest: Implications for Sampling

William W. Turechek and Walter F. Mahaffee

First author: Cornell University, New York State Agricultural Experiment Station, Department of Plant Pathology, Barton Laboratory, Geneva 14456; and second author: U.S. Department of Agriculture-Agriculture Research Service-Horticulture Crops Research Laboratory, Corvallis, OR 97330.

Accepted for publication 24 May 2004.

## ABSTRACT

Turechek, W. W., and Mahaffee, W. F. 2004. Spatial pattern analysis of hop powdery mildew in the Pacific Northwest: Implications for sampling. *Phytopathology* 94:1116-1128.

The spatial pattern of hop powdery mildew was characterized using 3 years of disease incidence data collected in commercial hop yards in the Pacific Northwest. Yards were selected randomly from yards with a history of powdery mildew, and two to five rows were selected for sampling within each yard. The proportion of symptomatic leaves out of 10 was determined from each of  $N$  sampling units in a row. The binomial and the beta-binomial frequency distributions were fit to the  $N$  sampling units observed in each row and to  $\Sigma N$  sampling units observed in each yard. Distributional analyses indicated that disease incidence was better characterized by the beta-binomial than the binomial distribution in 25 and 47% of the data sets at the row and yard scales, respectively, according to a log-likelihood ratio test. Median values of the beta-binomial parameter  $\theta$ , a measure of small-scale aggregation, were near 0 at both sampling scales, indicating that disease incidence was close to being randomly distributed. The variability in disease incidence among rows sampled in the same yard generally increased with mean incidence at the

yard scale. Spatial autocorrelation analysis, used to measure large-scale patterns of aggregation, indicated that disease incidence was not correlated between sampling units over several lag distances. Results of a covariance analysis showed that heterogeneity of disease incidence was not dependent upon cultivar, region, or time of year when sampling was conducted. A hierarchical analysis showed that disease incidence at the sampling unit scale (proportion of sampling units with one or more diseased leaves) increased as a saturation-type curve with respect to incidence at the leaf level and could be described by a binomial function modified to account for the effects of heterogeneity through an effective sample size. Use of these models permits sampling at the sampling unit scale while allowing inferences to be made at the leaf scale. Taken together, hop powdery mildew was nearly randomly distributed with no discernable foci, suggesting epidemics are initiated from a well-distributed or readily dispersible overwintering population. Implications for sampling are discussed.

*Additional keywords:* *Humulus lupulus*, nested analysis of variance, *Podosphaera macularis*, quantitative epidemiology, *Sphaerotheca humulus*.

Hop (*Humulus lupulus* L.) is a rhizomous climbing perennial plant grown commercially for its harvestable cones. The cones are used in the production of beer to provide it with its characteristic aromatic and flavoring properties. In commercial production, hop plants are grown on a 3 to 6.4 m high trellis system, depending upon cultivar and region cultivated. Hops are cultivated in several countries worldwide with major production in Australia, Brazil, the Czech Republic, Germany, and the United States. In the United States, the commercial hop industry is located in the Pacific Northwestern states of Washington, Oregon, and Idaho. Local production can be found in a number of other states such as California, Maine, New York, Vermont, and Wisconsin that generally supply hops to local breweries and brew pubs.

Hop powdery mildew, caused by *Podosphaera macularis* Braun & Takamatus (formerly *Sphaerotheca macularis* syn. *S. humuli*), is one of the oldest and most serious diseases of hop (26). It has long been a major problem in European hop production and was partially responsible for pushing the hop industry out of the eastern United States and California and into the Pacific Northwest (26,33). In June 1997, despite years of quarantine efforts, pow-

dery mildew was discovered in hop yards in Toppenish, WA (28). Immediately, the movement of hop plant material across the three Pacific Northwest states was restricted to prevent spread of disease to Oregon and Idaho. Unfortunately, powdery mildew was observed in both Oregon and northern and southern Idaho in 1998 and annually becomes epidemic in all growing regions.

The disease affects all aboveground parts of the plant; however, it is infections of the burrs (the female flower) and immature cones (the fruit) that result from secondary cycles of leaf infections that cause the most appreciable damage and reduction in crop yield and value (31). Infected burrs fail to develop into cones, and infection of developing cones typically results in deformity, discoloration (affecting quality), and/or reduction in weight. Leaves are susceptible to infection once they unfurl until they reach maturity, approximately 2 weeks after emergence (38) or earlier when exposed to supra-optimal temperature (25). The optimum temperature for foliar infection and disease development is approximately 18°C and the risk of infection decreases substantially once the temperature exceeds 30°C; however, infections can occur at temperatures as high as 36°C as long as exposure is for a short duration and it is followed by temperatures conducive for disease development (25).

Currently, well-timed fungicide applications are the only effective means to controlling powdery mildew on susceptible cultivars, and growers typically spend from \$300 to \$600 per hectare on fungicide applications to manage the disease. Because hops grow continuously from shoot emergence through harvest, the

Corresponding author: W. W. Turechek; E-mail address: wwt3@cornell.edu

Publication no. P-2004-0726-01R

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2004.

availability of susceptible tissue is not a limiting factor to epidemic development. Thus, growers or consultants must make management decisions based on the susceptibility of the cultivar, weather conditions, and the current level of infection in yards. The susceptibility of a cultivar is generally known prior to its commercial release and, recently, Mahaffee et al. (24) adapted the Gubler-Thomas grape powdery mildew forecaster (10) for hop to help growers select an appropriate fungicide schedule. This information has helped reduce the number of fungicide applications, but there is still economic and environmental need to further reduce the amount of fungicides used.

Despite the progress that has been made in disease management, there has been little advancement in our knowledge of the epidemiology of the disease in hop yards (26,31). This limits our ability to develop novel management practices and test hypotheses that specifically address hop powdery mildew. Spatial pattern is one of the fundamental epidemiological characteristics of a disease (3,20). Dispersal, establishment, growth, competition, and survival, for example, are inherently spatial processes (5,18,29,40). Although it is generally accepted that process cannot be inferred from pattern—the observed pattern at any given time is the result of any number of interacting past processes—it is widely believed that the observed pattern is a powerful tool for generating testable hypotheses (5). Moreover, knowledge of the spatial pattern of disease can be used explicitly in the design of sampling plans (2,34) assessment of crop loss (9,12,22), the analysis of designed experiments (7,14,23) and could be useful in the development of economical management practices and risk analysis (24).

There are many different approaches for characterizing the spatial pattern of disease. For disease incidence data, characterizing the frequency distribution of the proportion of diseased individuals per sampling unit is one widely used approach (4,6,20,32). Distributional analyses, as these are often referred to, provide a direct measurement of heterogeneity of disease incidence at the scale of the sampling unit. Distributional analyses do not require knowledge of the spatial location of the sampling units, only the number of diseased individuals within each sampling unit. The binary power law (13) is an extension of the distributional approach and can be used for the identification of factors that affect the heterogeneity of disease incidence. For example, it is unknown whether there exists any regional, varietal, or temporal variation in the spatial distribution of hop powdery mildew. These and other factors can be included in a linear model to quantify their effects on the heterogeneity of disease incidence (35,36). In contrast, correlation-type analyses are useful for characterizing the spatial relationship or association among sampling units (35,36). These analyses do require the relative location of the sampling units to be known and provide complimentary information to the distributional analyses.

Second, Turechek and Madden (37) recently introduced several models useful for characterizing the relationship of disease incidence in a spatial hierarchy. Essentially, use of these models permits sampling or the collection of data at one hierarchical scale—presumably the scale easiest to collect information or one less prone to sampling error—and allows one to make inferences about the incidence of disease at a more relevant scale (16). It is important to make the distinction between the more traditional use of the term scale and hierarchical scale; both are used here. The traditional use refers to a distinctive relative size such as the dimension of a sampling unit (e.g., 1 versus 4-m<sup>2</sup> sampling unit). Hierarchical scale refers to a series of scales or levels where each level of the series is a subset of the level above it. Hierarchical scales can be artificial, e.g., a yard can be divided into a grid of 1-m<sup>2</sup> quadrats and levels above are the result of combining quadrats, or they can be natural, e.g., leaves represent the lowest level, whereas the branch the leaves are attached to, the tree the branch is attached to, the orchards the tree is located in, etc., represent

successively higher levels of the hierarchy. The ability to make inferences from a scale that can be efficiently sampled to a scale more relevant for decision-making increases our disease management abilities.

The objectives of this research were to (i) characterize the spatial pattern (i.e., incidence and heterogeneity) of hop powdery mildew on foliage in hop yards in the Pacific Northwest and identify factors that impact spatial heterogeneity; and (ii) characterize the hierarchical relationship of hop powdery mildew between the leaf and sampling unit scales using recently introduced statistical models (37). The overall aim is to provide the information necessary to determine at which spatial and/or hierarchical scale sampling should be conducted to obtain a precise estimate of disease incidence, and whether the sampling strategy should be altered relative to cultivar, location, or time of year the sampling is conducted. Practically, an increased understanding of the spatial distribution of hop powdery mildew would further our ability to develop rational disease management practices and cost-effective sampling procedures for research and routine scouting.

## MATERIALS AND METHODS

Between 1999 and 2001, epidemics of powdery mildew were monitored in hop yards in Washington and Oregon selected from yards with a history of the disease. In Washington, yards from each of the major hop growing regions within the Yakima Valley were selected. From east to west, the regions are named based on the town central to the region; Prosser, Mabton, Moxee, and the Yakima Indian Reservation (Reservation). In Oregon, farms were located in towns of Mt. Angel, Silverton, St. Paul, and Woodburn. These regions span the geographical distribution of hop production in Oregon and Washington.

Hop yards were generally very large for horticultural crops, in which single yards ranged from 8 to 32 ha and consisted of 20 to 180 rows. The yards ranged in age from 3 to 37 years. The majority of yards sampled were planted with 1 m spacing between hills and 4.25 m spacing between rows, although some yards were planted at 2.13 by 2.13 m spacing. Yards were drip-, furrow-, or sprinkler-irrigated, with the majority being drip-irrigated. Yards were planted to a single cultivar and most sampling was conducted in yards planted with cvs. Columbus, Tomahawk, or Zeus (genetically indistinguishable, thus collectively called CTZ) in Washington; and cvs. Willamette, Cascade, Nugget, Galena, and Perle in Oregon. CTZ are very susceptible to the strains of powdery mildew present in the Pacific Northwest, Galena and Perle are moderately susceptible, Cascade is moderately resistant, and Nugget is considered resistant. Willamette is moderately resistant when grown in Washington and southern Idaho while highly susceptible in Oregon. This difference is due to the plants response to environmental conditions and not due to different pathogen races in the regions (W. F. Mahaffee, *unpublished data*).

**Data collection.** Foliar symptoms were assessed every 3 to 4 weeks in each yard beginning in May and continuing through harvest in late August or September. In 1999, a single transect, herein referred to as a row, was sampled from each yard. In 2000, each yard was partitioned into  $H$  strata, where  $H$  = (number of rows in a yard)/20 [rounded up to the nearest integer]. A single number,  $r$ , between 1 and 20 was randomly chosen and the  $r^{\text{th}}$  row from each stratum was sampled as described below. The partitioning of yards into strata permitted us to evaluate the variability among rows within individual yards and was done to determine if the results obtained from a single-row sample was representative of the disease status of the entire yard. In 2001, row selection was modified such that the first 40 rows from the north or eastern side of each yard (depending upon its orientation) were divided into 2 strata of 20 rows, and a single row within each stratum was randomly selected at each sampling date and sampled as in 2000. This was done to standardize the sampling

strategy across yards, and permitted us to include more yards in the survey in 2001.

Each yard was sampled multiple times over the course of the season, but transects were chosen randomly on each sampling date. In 1999, 44 transects from 14 yards were sampled (2 from OR, 12 from WA); in 2000, 370 transects were sampled from 42 yards (15 from OR, 27 from WA); and in 2001, 164 transects were sampled from 54 yards (20 from OR, 34 from WA). All yards sampled in 1999 were included in the 2000 sampling; all yards sampled in 2000 were included in the 2001 sampling.

The selection of sampling units within rows was identical each year. In each row,  $n = 10$  leaves were arbitrarily selected from each plant within the row or from the first  $N = 100$  (in 1999) or  $N = 75$  plants (in 2000 and 2001), whichever came first, and each leaf was rated for the presence or absence of powdery mildew. This is called cluster sampling (15) where each of  $N$  clusters (i.e., sampling units) contains  $n$  leaves. Thus,  $nN$  leaves were collected from each row and  $n\sum N_H$  leaves were collected from each yard, where  $N_H$  is the number of sampling units in the  $H^{\text{th}}$  stratum, where  $H = 1$  in 1999,  $H = 2$  in 2001, and  $H$  was variable in 2000.

**Distributional analyses.** The binomial and beta-binomial frequency distributions were used to characterize the spatial pattern of disease incidence (20). The binomial distribution has a single parameter,  $\pi$ , representing the probability of disease; in this case, the probability of having 0,1,2,...,10 diseased leaves in a sampling unit of  $n = 10$  leaves. The beta-binomial has two parameters,  $p$ , which is the expected probability of disease, and  $\theta$ , a measure of the variation, or heterogeneity, in disease incidence among sampling units above what would be expected from a good fit to the binomial distribution (i.e.,  $\theta = 0$ ). A good fit to the binomial distribution is an indication of a random spatial pattern of disease incidence while a good fit to the beta-binomial is indicative of an aggregated pattern.

The binomial and beta-binomial distributions were fit to the observed frequency data using the program BBD (19) for (i) the  $N$  sampling units observed in each individual row, herein referred to as "row-level analysis"; and (ii) the  $\sum N_H$  sampling units observed in each yard, herein referred to as "yard-level analysis". An estimate of disease incidence ( $= \hat{p}$ ) for a single row was obtained using  $\sum x/nN$ ; an estimate of disease incidence for a single yard ( $= \hat{p}_y$ ) was obtained using  $\sum x/n\sum N_H$ , where  $x$  is the number of diseased leaves in a single sampling unit of size  $n$  summed across all sampling units in a row or a yard for row- and yard-level analyses, respectively. The  $y$  subscript refers to measurements or parameters taken at the yard scale. The heterogeneity parameters,  $\theta$  and  $\theta_y$ , were obtained using the method of maximum likelihood as described in Madden and Hughes (19); these were also calculated with BBD. When it was not possible to obtain maximum likelihood estimates of the binomial or beta-binomial parameters, the moment estimates were used in all subsequent calculations (20).

Chi-square goodness-of-fit tests were calculated for each distribution after disease incidence classes (i.e., 0 to 10 diseased leaves) were pooled so expected frequencies in each class exceeded five. Because a goodness-of-fit test cannot be performed when the number of incidence classes is equal to or is less than the degrees of freedom for the test, a log-likelihood ratio test statistic was also calculated to test whether the beta-binomial distribution fit the data better than the binomial distribution (34). Additionally, the  $C(\alpha)$  test was used to test specifically whether heterogeneity could be described by the beta-binomial distribution.

The degree of aggregation was estimated using the  $\theta$  parameter of the beta-binomial distribution and the index of dispersion,  $D$  (20). When the distribution of disease incidence is random,  $\theta = 0$  and  $D = 1$ ; when disease incidence is aggregated, the observed variance is greater than the expected binomial variance making  $\theta > 0$  and  $D > 1$ .

**Power law analyses.** The binary power law was expressed as the relationship between the observed sample variance of diseased leaves ( $v_{\text{obs}}$ ) and the theoretical variance of the binomial distribution [ $v_{\text{bin}} = n\hat{p}(1 - \hat{p})$ ],

$$\ln(v_{\text{obs}}) = \ln(A_x) + b \ln[n\hat{p}(1 - \hat{p})] \quad (1)$$

where  $\ln(A_x)$  and  $b$  are the intercept and slope of a straight line, respectively, and  $\hat{p}$  is the moment estimate of  $p$  from either the binomial or beta-binomial distribution. When  $A_x = 1$  and  $b = 1$ , diseased leaves have a random spatial pattern that can be described by the binomial distribution. When  $b = 1$  and  $A_x > 1$ , diseased leaves have an aggregated distribution but the degree of aggregation does not depend on  $p$ . When  $b$  and  $A_x$  are both greater than 1, the degree of aggregation or heterogeneity changes systematically with  $p$ . If the data can be described by the beta-binomial distribution, then there is a direct mathematical relationship between  $\theta$  of the beta-binomial distribution and the parameters of the binary power law:

$$\theta = [a - f(p)/n]/[f(p) - a] \quad (2)$$

where  $f(p) = [p(1 - p)]^{1-b}$  and  $a = A_x n^{b-2}$  (20,35). Equation 2 represents a curve with maximum value at  $p = 0.5$  when  $b > 1$ , and a horizontal line when  $b = 1$ . The relationship between  $\theta$  and  $p$  is important for characterizing the spatial process and for sampling where estimates of  $\theta$  are needed to develop fixed and sequential sampling plans (34,36).

**Covariance analysis.** A covariance analysis was performed to determine the effect of the factors year (first, second, and third year), region (Oregon, Prosser, Mabton, Reservation, and Moxee), assessment date (categorized by month), and cultivar (CTZ, Willamette, Perle, and the remainder in the final grouping) on the slope ( $b$ ) and intercept [ $\ln(A_x)$ ] of the power law (equation 1). A covariance analysis was done for each year separately, and then a separate covariance analysis was conducted to determine if year affected estimates of  $\ln(A_x)$  or  $b$ .

To perform the analysis, the power law model (equation 1) was considered the null model and factors were added individually, first as an intercept term, then as an interaction term with the slope. A factor was considered significant if the new model had a deviance significantly lower than the original power law model. The test of a factor is not an indication of whether the factor had an influence on disease incidence or heterogeneity, but whether the factor affected the degree of heterogeneity after first correcting for the relationship between heterogeneity and  $v_{\text{ran}}$  (35). The covariance analysis was performed using GLIM (Generalized Linear Interactive Modeling, version 3.77, Royal Statistical Society, Oxford).

**Autocorrelation analysis.** First- and second-order autocorrelation coefficients were calculated for each individual row to determine the degree of correlation between neighboring sampling units (35). Data were transformed using the logit transformation,  $\ln[y/(1 - y)]$ , where  $y = (x + 0.5)/(n + 1)$  is the Haldane transformation of  $x$ , and  $x$  is the number of diseased leaflets in each sampling unit. Autocorrelation coefficients were not calculated at the yard level because the spatial proximity of sampling units in different rows was not calculated.

**Incidence of sampling units.** The zero term [ $\text{Pr}(0)$ ] and mean ( $p$  or  $p_y$ ) of the probability distribution of diseased leaves was used to estimate the proportion of diseased sampling units in a row ( $p_{su}$ ) or a yard ( $p_{y,su}$ ) (37). If  $\text{Pr}(0)$  represents the probability that all leaves in a sampling unit were disease-free, then  $p_{su} = 1 - \text{Pr}(0)$  represents the probability that at least one leaf was diseased. For simplicity in presentation, we will refer only to the relationship at the row level. Derivation at the yard scale follows exactly except that yard-level statistics replace row-level statistics.

If the incidence of diseased leaves can be described by a binomial distribution, then

$$\tilde{p}_{su} = 1 - (1 - \hat{p})^n \quad (3)$$

A tilde (~) is used to denote that the estimate of  $p_{su}$  was derived from measurements at another spatial scale. If incidence of diseased leaves can be described by the beta-binomial distribution, then

$$\tilde{p}_{su} = 1 - \prod_{j=0}^{n-1} \frac{1 - \hat{p} + j\hat{\theta}}{1 + j\hat{\theta}} \quad (4)$$

where  $\hat{\theta}$  is described by equation 2. Equation 4 is difficult to rearrange to allow prediction of  $p$  from  $p_{su}$ . However, equation 3 can be modified to (empirically) correct for the effect of heterogeneity or, in some cases, serve as an approximation to equation 4. One approach is to follow Rao and Scott (30) and replace  $n$  in equation 3 with an “effective sample size”. The effective sample size represents the reduction in information obtained in the sample due to the effects of heterogeneity. There are several ways to derive an effective sample size; three were considered here. The first, and most common, effective sample size was derived by dividing  $n$  by the index of dispersion ( $D$ ) (17),

$$n_D = n/D \quad (5)$$

Since there were several data sets, the average value of  $D$  was used. Doing this, however, relies on the assumption that  $D$  is not systematically related to  $p$ . Since  $b > 1$  in the binary power law (equation 1), indicating that  $D$  is systematically related to  $p$ , a second effective sample size based on the parameters of the power law was derived using

$$n_{\hat{p}} = f(\hat{p})/\hat{a} \quad (6)$$

where  $f(\hat{p})$  and  $\hat{a}$  are defined above. Madden and Hughes (21) developed an iterative and graphic method to derive an effective sample size for data that are beta-binomially distributed. This third effective sample size was calculated using

$$n_{\hat{\theta}} = n/(0.98135 + 0.8179\hat{\theta} + 0.004958n + 0.30387\hat{\theta}n - 0.3471\hat{\theta}^2 - 0.08475\hat{\theta}^2n) \quad (7)$$

where  $\hat{\theta}$  is the heterogeneity parameter of the beta-binomial distribution estimated using the method of maximum likelihood or with equation 2 as described previously.

To determine which of these effective sample size models was the best substitute for  $n$  in equation 3 to describe the relationship between  $\hat{p}$  and  $\hat{p}_{su}$ , a generalized linear modeling approach, as described by Farrington (8) and Turechek and Madden (37), was taken. A brief explanation of the approach is given in the Appendix.

**Estimation of variance components.** A fully nested analysis of variance (ANOVA) was used to estimate the variance compo-

nents at each scale of sampling for the incidence of powdery mildew (2). The analysis was conducted such that sampling units were nested within rows, rows were nested within yards, and yards were nested within region. The analysis was performed only with data collected in 2000 because these represented the yards that were sampled most intensively. Data were transformed using the arcsine-square root transformation of  $p$ , prior to the analysis.

## RESULTS

**Distributional analyses. Row level.** Disease incidence varied by year and location (Table 1). Over the 3 years of the study and across all regions, the distribution of disease incidence was positively skewed with values ranging from 0.00 to 0.77 and a median of 0.012 (Fig. 1A). For all 3 years of the study, disease incidence was highest in yards located in the Reservation and lowest in yards located in Oregon. The heterogeneity parameter of the beta-binomial distribution ( $\hat{\theta}$ ) ranged from 0.00 to 0.76 with a median of 0.006, indicating a low degree of aggregation (Fig. 1C). As expected based on equation 2,  $\hat{\theta}$  was slightly higher for yards located in the Reservation, where  $\hat{p}$  was highest, and lowest for yards located in Oregon (Table 2; Fig. 2A). However, median values of  $\hat{\theta}$  were higher than expected in the highest incidence class, where values are expected to decrease after achieving a maximum at  $\hat{p} = 0.5$  according to equation 2.

Maximum likelihood estimation of the beta-binomial parameters was possible for 300 (52%) data sets at the row level (Table 3). For the 278 cases where maximum likelihood estimation was not possible, the moment estimates of  $\hat{p}$  were  $<0.056$ —the maximum  $\hat{p}$  of the 278 cases—or the moment estimate of  $\hat{\theta}$  was 0. Of the 300 data sets where maximum likelihood estimation of  $\hat{\theta}$  was possible,  $\hat{\theta}$  was significantly greater than 0 in 74 (25%) data sets. The index of dispersion,  $D$ , ranged from 0.4 to 4.94 with mean 1.25 and median 1.07 (Fig. 1E).  $D$  was significantly greater than 1 in 33% of the data sets according to a chi-square test. Similarly, the  $C(\alpha)$  test indicated that the beta-binomial distribution fit the data in 35% of the data sets.

The beta-binomial distribution fit the data better than the binomial distribution in 25% of the 578 data sets according to a log likelihood ratio test; as incidence increased, the data were more likely to be described by the beta-binomial distribution (Table 3). Where there were sufficient numbers of classes to allow the chi-square goodness-of-fit test to be performed, 107/121 could be described by the beta-binomial distribution and 84/164 could be described by the binomial distribution; 55 of these data sets could be described by both distributions. (More data sets could be tested for fit to the binomial distribution because one less degree of freedom was needed to perform the test.)

**Yard level.** Similar to values collected at the row level, disease incidence was positively skewed with values ranging from 0.0004 to 0.71 with median 0.009 over the 3 years of the study

TABLE 1. Mean and standard error (SE) of the mean of the incidence of hop powdery mildew ( $\hat{p}$ ) in hop yards located in Washington and Oregon in 1999, 2000, and 2001

| Date and scale <sup>a</sup> | Oregon                    |       |  | Prosser                   |     |  | Mabton                    |     |  | Reservation               |     |  | Moxee                     |     |  | Combined                  |     |  |
|-----------------------------|---------------------------|-------|--|---------------------------|-----|--|---------------------------|-----|--|---------------------------|-----|--|---------------------------|-----|--|---------------------------|-----|--|
|                             | $\hat{p}$ (SE $\hat{p}$ ) | $T^b$ |  | $\hat{p}$ (SE $\hat{p}$ ) | $T$ |  | $\hat{p}$ (SE $\hat{p}$ ) | $T$ |  | $\hat{p}$ (SE $\hat{p}$ ) | $T$ |  | $\hat{p}$ (SE $\hat{p}$ ) | $T$ |  | $\hat{p}$ (SE $\hat{p}$ ) | $T$ |  |
| 1999 row                    | 0.10 (0.048)              | 7     |  | 0.180 (0.050)             | 14  |  | 0.08 (0.030)              | 18  |  | 0.49 (0.074)              | 5   |  | —                         | —   |  | 0.16 (0.029)              | 44  |  |
| 1999 yard                   | —                         | —     |  | —                         | —   |  | —                         | —   |  | —                         | —   |  | —                         | —   |  | —                         | —   |  |
| 2000 row                    | 0.01 (0.001)              | 157   |  | 0.006 (0.001)             | 53  |  | 0.03 (0.006)              | 50  |  | 0.12 (0.018)              | 77  |  | 0.09 (0.023)              | 33  |  | 0.05 (0.005)              | 370 |  |
| 2000 yard                   | 0.01 (0.002)              | 38    |  | 0.004 (0.001)             | 22  |  | 0.03 (0.008)              | 19  |  | 0.12 (0.035)              | 20  |  | 0.06 (0.026)              | 20  |  | 0.04 (0.008)              | 119 |  |
| 2001 row                    | 0.02 (0.003)              | 61    |  | 0.007 (0.003)             | 4   |  | 0.13 (0.026)              | 20  |  | 0.15 (0.022)              | 54  |  | 0.09 (0.021)              | 25  |  | 0.08 (0.010)              | 164 |  |
| 2001 yard                   | 0.01 (0.003)              | 35    |  | 0.007 (0.004)             | 2   |  | 0.12 (0.037)              | 10  |  | 0.15 (0.044)              | 19  |  | 0.08 (0.027)              | 13  |  | 0.07 (0.014)              | 79  |  |
| All row                     | 0.02 (0.002)              | 225   |  | 0.040 (0.013)             | 71  |  | 0.07 (0.010)              | 88  |  | 0.15 (0.015)              | 136 |  | 0.09 (0.016)              | 58  |  | 0.07 (0.005)              | 578 |  |
| All yard                    | 0.01 (0.002)              | 73    |  | 0.004 (0.001)             | 24  |  | 0.06 (0.019)              | 29  |  | 0.14 (0.028)              | 39  |  | 0.07 (0.019)              | 33  |  | 0.05 (0.008)              | 198 |  |

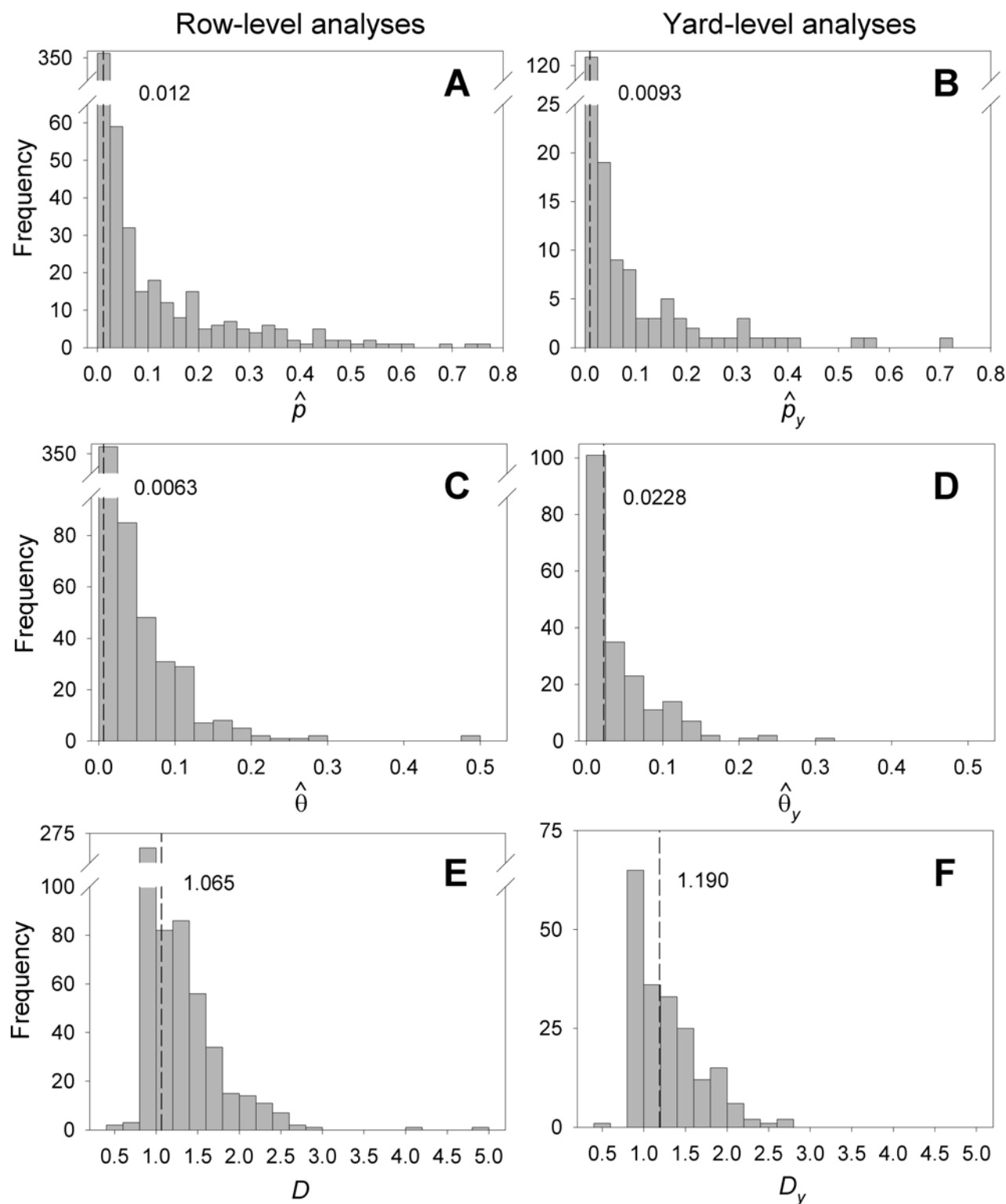
<sup>a</sup> Year of sampling and scale of assessment.

<sup>b</sup> Number of data sets.

and across all regions (Table 1; Fig. 1B). The heterogeneity parameter,  $\hat{\theta}_y$ , was also positively skewed with values ranging from 0.00 to 0.31 and a median value of 0.022, indicating a low degree of aggregation (Table 1; Fig. 1D). Similar to row-level data,  $\hat{\theta}_y$  increased with  $\hat{p}_y$  as expected according to equation 2, but again, values of  $\hat{\theta}_y$  were highly variable and, on average, slightly higher than expected at higher levels of incidence (Fig.

2B). Note, however, that most incidence values were below 0.5 and that the maximum value of incidence was 0.71, thus providing little data to “fill” the upper portion of the prediction curve (equation 2).

Maximum likelihood estimation was possible for 124 (63%) data sets at the yard level (Table 3). For the 73 cases where maximum likelihood estimation was not possible, the moment esti-



**Fig. 1.** Frequency distributions of the estimated beta-binomial parameters  $\hat{p}$  and  $\hat{\theta}$ , and the index of dispersion ( $D$ ) for incidence of hop powdery mildew on leaves for data collected at the **A, C, and E**, row level or **B, D, and F**, yard level. The frequency distributions are based on 578 assessments at the row level and 198 assessments at the yard level using data collected in Oregon and Washington from 1999 to 2001. Vertical dashed lines represent the median value for that statistic; medians are given numerically on the graph.

mates of  $\hat{p}_y$  were either  $<0.057$  or the moment estimate of  $\hat{\theta}_y$  was 0. Of the 124 data sets where maximum likelihood estimation of  $\hat{\theta}_y$  was possible,  $\hat{\theta}_y$  was significantly greater than 0 in 67 (54%) data sets. The index of dispersion,  $D_y$ , ranged from 0.58 to 2.68 with mean 1.31 and median 1.19 (Fig. 1F).  $D_y$  was significantly greater than 1 in 53% of the data sets according to a chi-square test. The  $C(\alpha)$  test indicated that the beta-binomial distribution fit the data in 56% of the data sets.

The beta-binomial distribution fit the data better than the binomial distribution in 47% of the 198 data sets according to a log likelihood ratio test (Table 3). Where there were sufficient classes to allow the chi-square goodness-of-fit test to be performed, 48/65 could be described by the beta-binomial distribution and 8/70 could be described by the binomial distribution; three of these data sets could be described by both distributions.

TABLE 2. Mean value of the beta-binomial parameter  $\theta$  and its standard error (SE) for hop powdery mildew in hop yards located in Washington and Oregon in 1999, 2000, and 2001

| Date and scale <sup>a</sup> | Oregon                              |       | Prosser                             |     | Mabton                              |     | Reservation                         |     | Moxee                               |     | Combined                            |     |
|-----------------------------|-------------------------------------|-------|-------------------------------------|-----|-------------------------------------|-----|-------------------------------------|-----|-------------------------------------|-----|-------------------------------------|-----|
|                             | $\hat{\theta}$ (SE $\hat{\theta}$ ) | $T^b$ | $\hat{\theta}$ (SE $\hat{\theta}$ ) | $T$ | $\hat{\theta}$ (SE $\hat{\theta}$ ) | $T$ | $\hat{\theta}$ (SE $\hat{\theta}$ ) | $T$ | $\hat{\theta}$ (SE $\hat{\theta}$ ) | $T$ | $\hat{\theta}$ (SE $\hat{\theta}$ ) | $T$ |
| 1999 row                    | 0.03 (0.014)                        | 7     | 0.06 (0.017)                        | 14  | 0.03 (0.012)                        | 18  | 0.04 (0.014)                        | 5   | —                                   | —   | 0.04 (0.008)                        | 44  |
| 1999 yard                   | —                                   | —     | —                                   | —   | —                                   | —   | —                                   | —   | —                                   | —   | —                                   | —   |
| 2000 row                    | 0.02 (0.003)                        | 157   | 0.01 (0.003)                        | 53  | 0.02 (0.005)                        | 50  | 0.02 (0.004)                        | 77  | 0.02 (0.005)                        | 33  | 0.02 (0.002)                        | 370 |
| 2000 yard                   | 0.03 (0.005)                        | 38    | 0.01 (0.003)                        | 22  | 0.04 (0.006)                        | 19  | 0.05 (0.013)                        | 20  | 0.02 (0.006)                        | 20  | 0.03 (0.003)                        | 119 |
| 2001 row                    | 0.04 (0.013)                        | 61    | 0.02 (0.013)                        | 4   | 0.06 (0.008)                        | 20  | 0.08 (0.012)                        | 54  | 0.06 (0.015)                        | 25  | 0.06 (0.007)                        | 164 |
| 2001 yard                   | 0.04 (0.008)                        | 35    | 0.02 (0.024)                        | 2   | 0.05 (0.011)                        | 10  | 0.08 (0.010)                        | 19  | 0.07 (0.025)                        | 13  | 0.05 (0.006)                        | 79  |
| All row                     | 0.02 (0.004)                        | 225   | 0.02 (0.005)                        | 71  | 0.03 (0.004)                        | 88  | 0.05 (0.006)                        | 136 | 0.04 (0.007)                        | 58  | 0.03 (0.002)                        | 578 |
| All yard                    | 0.03 (0.005)                        | 73    | 0.01 (0.003)                        | 24  | 0.04 (0.006)                        | 29  | 0.06 (0.008)                        | 39  | 0.04 (0.011)                        | 33  | 0.04 (0.003)                        | 198 |

<sup>a</sup> Year of sampling and scale of assessment.  
<sup>b</sup> Number of data sets.

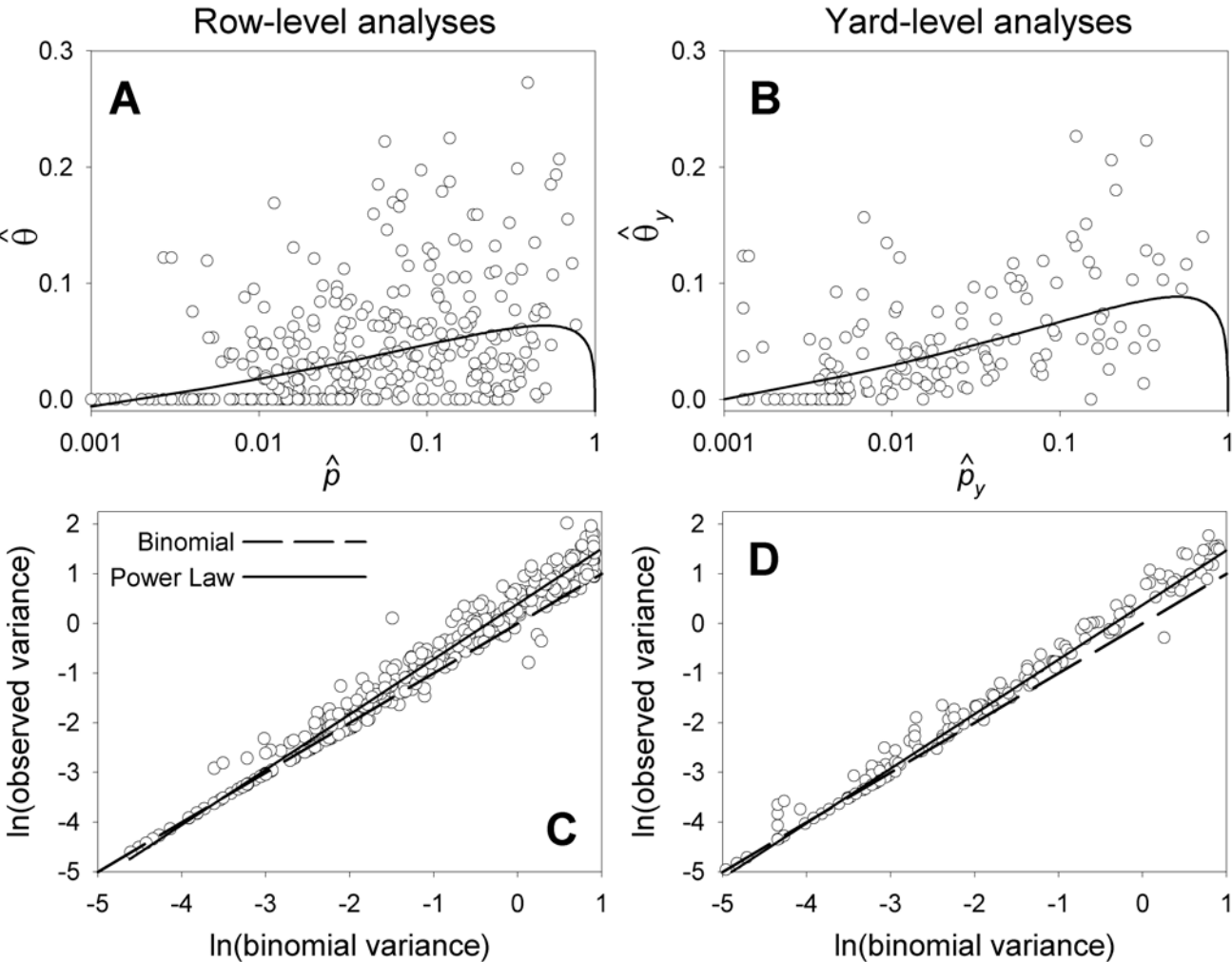


Fig. 2. **A and B**, Relationship between  $\hat{\theta}$  of the beta-binomial distribution and disease incidence  $\hat{p}$  and **C and D**, relationship between the logarithm of the observed variance and logarithm of the theoretical variance for a binomial distribution,  $n\hat{p}(1-\hat{p})$ , for incidence of hop powdery mildew in Oregon and Washington from 1999 to 2001. Observed data are shown as open symbols and represent **A**, single-row or **B**, yard assessments of hop powdery mildew. Prediction curves for  $\hat{\theta}$  are based on equation 2. **C and D**, The solid line represents the least-squares fit to the data with parameters given in Table 4 (1999 to 2001), while the broken line represents the binomial line ( $\hat{\theta}=0$ ). **A and C** are based on 578 assessments at the row level, **B and D** are based on 198 assessments at the yard level.

**Power law.** The binary power law provided an excellent fit to the data for each of the 3 years of the survey for both row-level and yard-level analyses (Table 4; Fig. 2C and D). Estimated slopes and intercept parameters were significantly greater than 0 and 1 according to a *t* test, respectively, indicating that heterogeneity (i.e.,  $\theta$ ) changed systematically with disease incidence at both scales. The covariance analysis indicated that the factors year, region, assessment date, and cultivar had no effect on the parameter estimates for either row- or yard-level analyses when model comparisons were based on their deviances (residual sum

of squares) (Table 5). Consequently, the data were combined and the power law parameters re-estimated for the 3 years of row-level data and the 2 years of yard-level data. The new slope and intercept parameters for either of the combined data sets were not significantly different from their corresponding parameter estimates calculated for the individual years. Sequential *F* tests indicated that the factor year had a significant effect when included as an intercept term in the row-level analyses; that is, adjusting for year affected the height of the line of the power law (Table 5). Region also had a marginally significant effect on the intercept at the

TABLE 3. Results of spatial analyses characterizing the degree of aggregation of disease incidence of hop powdery mildew in Washington and Oregon in 1999, 2000, and 2001

| Incidence<br>class <sup>a</sup> | Row   |                  |                  |               |                  |         | Yard |      |      |               |                  |
|---------------------------------|-------|------------------|------------------|---------------|------------------|---------|------|------|------|---------------|------------------|
|                                 | $T^b$ | MLE <sup>c</sup> | LRS <sup>d</sup> | Median values |                  |         | $T$  | MLE  | LRS  | Median values |                  |
|                                 |       |                  |                  | $D^e$         | $\hat{\theta}^f$ | $r_1^g$ |      |      |      | $D_y$         | $\hat{\theta}_y$ |
| 0.00–0.05                       | 419   | 36.5             | 11.5             | 1.00          | 0.000            | −0.020  | 152  | 52.6 | 32.9 | 1.09          | 0.009            |
| 0.05–0.10                       | 47    | 91.5             | 64.4             | 1.44          | 0.049            | 0.068   | 17   | 94.1 | 94.1 | 1.58          | 0.068            |
| 0.10–0.20                       | 53    | 90.5             | 64.1             | 1.46          | 0.052            | 0.181   | 13   | 92.3 | 92.3 | 1.63          | 0.074            |
| 0.20–0.40                       | 40    | 92.5             | 42.1             | 1.32          | 0.035            | 0.106   | 12   | 100  | 91.7 | 1.69          | 0.083            |
| 0.40–0.60                       | 15    | 100              | 73.3             | 1.62          | 0.072            | 0.155   | 3    | 100  | 100  | 1.85          | 0.101            |
| 0.60–0.80                       | 4     | 100              | 100              | 2.11          | 0.136            | 0.038   | 1    | 100  | 100  | 2.12          | 0.140            |

<sup>a</sup> Classes end with the listed incidence value and begin with the first value above the listed value in each class.

<sup>b</sup> Number of data sets in defined incidence class.

<sup>c</sup> Percentage of data sets in which the maximum likelihood estimation (MLE) procedure converged to provide estimates of the beta-binomial parameters *p* and  $\theta$  (19).

<sup>d</sup> Percentage of data sets in which the beta-binomial distribution fit the data better than the binomial distribution according to the likelihood ratio test.

<sup>e</sup> Median index of dispersion.

<sup>f</sup> Median estimated beta-binomial parameter.

<sup>g</sup> Median first-order autocorrelation statistic.

TABLE 4. Estimated parameters of the binary power law (equation 1) and their standard errors (SE) for hop powdery mildew in hop yards located in Washington and Oregon in 1999, 2000, and 2001<sup>a</sup>

|                     | Year      | df  | <i>r</i> <sup>2</sup> | ln( <i>A</i> <sub><i>x</i></sub> ) | SE ln( <i>A</i> <sub><i>x</i></sub> ) | <i>b</i> | SE <i>b</i> |
|---------------------|-----------|-----|-----------------------|------------------------------------|---------------------------------------|----------|-------------|
| Row-level analysis  | 1999      | 43  | 98.6                  | 0.341                              | 0.037                                 | 1.104    | 0.020       |
|                     | 2000      | 369 | 98.5                  | 0.266                              | 0.019                                 | 1.063    | 0.007       |
|                     | 2001      | 163 | 97.7                  | 0.458                              | 0.030                                 | 1.112    | 0.013       |
|                     | 1999–2001 | 577 | 98.3                  | 0.350                              | 0.015                                 | 1.088    | 0.006       |
|                     |           |     |                       |                                    |                                       |          |             |
| Yard-level analysis | 2000      | 118 | 99.3                  | 0.417                              | 0.027                                 | 1.092    | 0.009       |
|                     | 2001      | 78  | 98.2                  | 0.490                              | 0.041                                 | 1.100    | 0.017       |
|                     | 2000–2001 | 197 | 98.9                  | 0.457                              | 0.023                                 | 1.099    | 0.008       |
|                     |           |     |                       |                                    |                                       |          |             |

<sup>a</sup> ln(*A*<sub>*x*</sub>) and *b* are the intercept and slope estimates, respectively, of the binary power law (equation 1).

TABLE 5. Deviances (Dev.) and sequential *F* tests for the covariance models testing for the effects of the factors year, season, region, cultivar, and yard on the intercept and slope parameters of the binary power law (equation 1) for hop powdery mildew in hop yards located in Washington and Oregon in 1999, 2000, and 2001

|            |           | Intercept <sup>b</sup>       |                              |        |                   |                |       | Slope <sup>b</sup> |        |       |       |
|------------|-----------|------------------------------|------------------------------|--------|-------------------|----------------|-------|--------------------|--------|-------|-------|
|            |           | df <sub>1</sub> <sup>a</sup> | df <sub>2</sub> <sup>a</sup> | Dev.   | Diff <sup>c</sup> | F <sup>d</sup> | P     | Dev.               | Diff   | F     | P     |
| Row-level  | Power law | 576                          |                              | 30.852 | —                 | —              | —     | 30.852             | —      | —     | —     |
|            | +Year     | 574                          | 2                            | 29.794 | 1.058             | 10.19          | 0.000 | 30.743             | 0.109  | 1.018 | 0.362 |
|            | +Season   | 573                          | 3                            | 30.795 | 0.057             | 0.354          | 0.787 | 30.845             | 0.007  | 0.043 | 0.988 |
|            | +Region   | 572                          | 4                            | 30.714 | 0.138             | 0.643          | 0.632 | 30.814             | 0.038  | 0.176 | 0.951 |
|            | +Cultivar | 573                          | 3                            | 30.767 | 0.085             | 0.528          | 0.663 | 30.830             | 0.022  | 0.136 | 0.938 |
|            | +Yard     | 524                          | 52                           | 27.765 | 3.087             | 1.120          | 0.269 | 28.786             | 2.066  | 0.723 | 0.926 |
|            |           |                              |                              |        |                   |                |       |                    |        |       |       |
| Yard-level | Power law | 196                          |                              | 7.8779 | —                 | —              | —     | 7.8779             | —      | —     | —     |
|            | +Year     | 195                          | 1                            | 7.7400 | 0.1379            | 3.474          | 0.064 | 7.8255             | 0.0524 | 1.306 | 0.255 |
|            | +Season   | 194                          | 2                            | 7.8527 | 0.0252            | 0.311          | 0.733 | 7.8528             | 0.0251 | 0.310 | 0.734 |
|            | +Region   | 192                          | 4                            | 7.4975 | 0.3804            | 2.435          | 0.049 | 7.5635             | 0.3144 | 1.995 | 0.097 |
|            | +Cultivar | 193                          | 3                            | 7.6235 | 0.2544            | 2.147          | 0.096 | 7.6476             | 0.2303 | 1.937 | 0.125 |
|            | +Yard     | 148                          | 48                           | 6.1107 | 1.7672            | 0.892          | 0.671 | 6.1758             | 1.7021 | 0.850 | 0.739 |
|            |           |                              |                              |        |                   |                |       |                    |        |       |       |

<sup>a</sup> df<sub>1</sub> = degrees of freedom for the model deviance; df<sub>2</sub> = degrees of freedom for the factor in test.

<sup>b</sup> ln(*A*<sub>*x*</sub>) and *b* are the intercept and slope parameters, respectively, of the binary power law (equation 1).

<sup>c</sup> The difference between the deviance for the power law model and the power law model with the factor included as either an intercept term or as an interaction with the slope term. If the deviance was reduced by  $\chi^2_{df,2}$ , the factor was considered to improve the model significantly.

<sup>d</sup> The *F* test is a test of a factors significance based on its sequential sum of squares. The significance of a factor under this test is not an indication that the factor improves the overall fit to the model and, consequently, does not serve as a basis for inclusion in the final model.

yard-level. However, the inclusion of year or region in the power law model did not reduce the deviance significantly relative to the power law model (i.e., the null model).

**Autocorrelation analysis.** First-order autocorrelation statistics ( $\hat{\rho}_1$ ) were slightly skewed, with values ranging from  $-0.28$  to  $0.73$  with a mean value of  $0.053$  and a median of  $-0.013$ . For the 3 years of data, 19.7% of the data sets had significant positive first-order correlations and only 3.6% of the data sets had significant second-order correlations ( $\hat{\rho}_2$ ) based on the values exceeding the upper bound of their approximate 95% 1-sided confidence interval (calculated using  $1.645/\sqrt{N}$ ). The results indicate a low degree of spatial correlation among neighboring plants. First-order correlations were not systematically related with incidence at either the row or sampling unit level (Table 3).

**Incidence of sampling units.** Figure 3A and B show the hierarchical relationship between incidence of diseased leaves ( $\hat{p}$  or  $\hat{p}_y$ ) and incidence of diseased sampling units ( $\hat{p}_{su}$  or  $\hat{p}_{y,su}$ ) for data collected at row and yard levels, respectively (16,37). The curves shown are the binomial model (case i, solid line), the beta-binomial model (case ii, broken line), effective sample size model based on the power law parameters (case iii, dash-dot line) and effective sample size model based on Madden and Hughes (21) (case iv, dash-dot-dot line), and the best fitting model with slope and intercept parameters estimated (case v, dotted line). No parameter estimation was necessary for i thru iv cases.

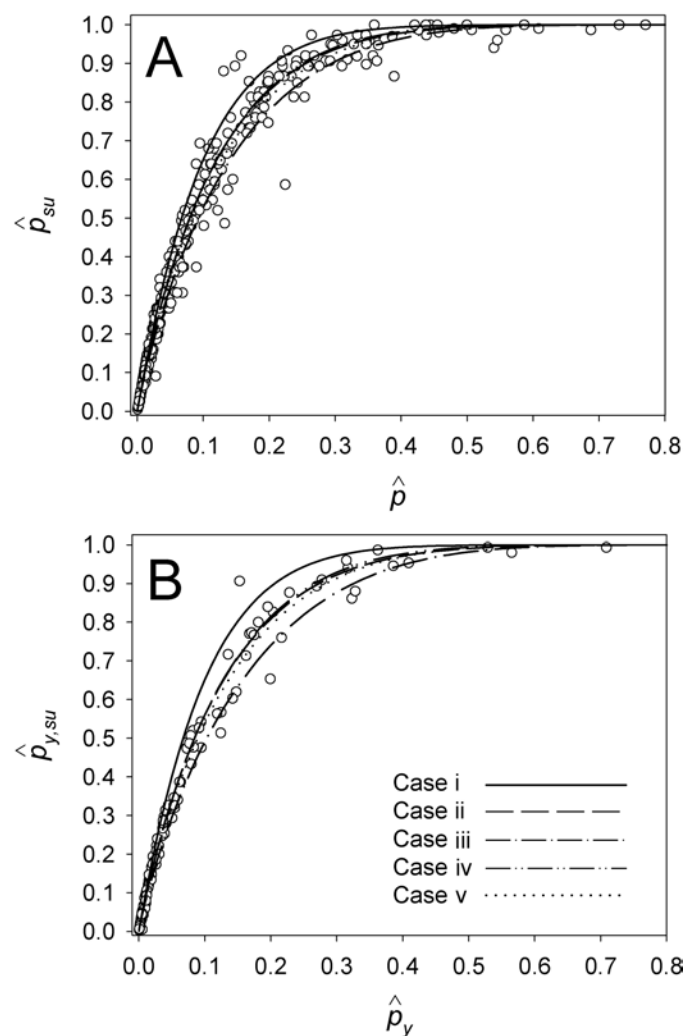
The binomial model provided the worst fit to both the row- and yard-level data (Table 6). Applying any correction for heterogeneity to the binomial model improved the fit to the data. Of the models that did not involve curve fitting, the effective sample size model based on Madden and Hughes (21) gave the best fit to the row-level data. The beta-binomial model provided the best fit to the yard-level data. The effective sample size model based on Madden and Hughes (21) closely approximated the beta-binomial model as expected. The best-fitting model, by definition, was obtained by fitting both the intercept and slope parameters via regression. These results were very similar to those obtained with strawberry leaf blight (37).

**Estimation of variance components.** ANOVA results indicated that the largest component of variation in incidence was attributed to variability among yards followed by sampling units, region, and then row (Table 7). Approximately 30% of the total variation in incidence could be explained by the variability among sampling units within rows. Relative to the variation among rows, this accounts for nearly 80% of the variation within a yard. Despite the magnitude of difference, the variation among rows within yards must be taken into account to minimize sampling error. Approximately 18% of the variation could be attributed to regional effects; this is seen in Figure 4 where the range of incidence values for single rows sampled in individual yards at the leaf (Fig. 4A) and sampling unit (Fig. 4C) scales are shown for yards sampled in 2000. (Eleven additional yards are represented in Figure 4 that are not listed in Tables 1 to 3 because yards with no powdery mildew are included). The overall higher incidence in the Reservation and Moxee regions is likely attributed to the density to which hops is planted in these regions, the fact that the Reservation was the region where the first epidemics were recorded, and less likely to regional variation in climate.

In contrast to incidence, values of  $\hat{\theta}$  were more variable within and among yards (Fig. 4B). This variability did not appear to be related to region or to incidence, supporting results of the covariance analysis.

**Sampling a yard.** The information in Figure 4 was used to calculate the probability of incorrectly concluding that the incidence of powdery mildew is less than an arbitrary incidence threshold relative to the number of rows sampled in a yard. Consider a threshold of  $p = 0.10$ , represented by the broken line in Figure 4A; the broken line in Figure 4B is the corresponding value of  $p_{su}$  estimated using equation A2 with the intercept and

slope parameters estimated using the best-fitting model, i.e., case v. In the context of sampling, we impose the rule that if any row sampled in a yard exceeds the threshold value, then an "action" is warranted. If the decision to act were to be based on sampling a single row, then over the 130 data sets, there are eight yards where the decision to act could have gone either way. That is, there are eight yards where individual rows had incidence values above and below the threshold value. Sampling two rows reduces the chances of reaching an incorrect decision to only three yards. If we sampled three rows, then all yards would have been classified correctly based on the rules developed. The same conclusions are reached irrespective of whether incidence was rated at the leaf level or sampling unit level for this threshold.



**Fig. 3.** Relationship between incidence of hop powdery mildew at the leaf and sampling unit scales for data collected at the **A**, row or **B**, yard level. Observed data are shown as open symbols and represent **A**, single-row or **B**, yard assessments of hop powdery mildew. Case i was derived by assuming that incidence could be described by the binomial distribution at the leaf scale (equation 3). Case ii was derived by assuming that incidence could be described by the beta-binomial distribution at the leaf scale (equation 4). Case iii was derived by assuming that incidence could be described by the beta-binomial distribution but that heterogeneity varies systematically with  $p$  according to the binary power law (equation A1 with  $n_s = n_g$ ). Binary power law parameters are given in Table 4 (1999 to 2001). Case iv was derived by assuming that incidence could be described by the beta-binomial distribution but with heterogeneity varying according to the effective sample size of Madden and Hughes (21) (equation A1 with  $n_s = n_v$ ). Case v was derived assuming that heterogeneity varied according to the power law with parameters estimated by fitting equation A2. Analyses are based on 578 assessments at the row level and on 198 assessments at the yard level for data collected between 1999 and 2001.



This process was repeated for threshold values of 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.5, and 0.75 for leaf sampling and the corresponding values at the sampling unit scale were calculated as described above (Fig. 5). The probability of making an incorrect decision is greatest at low threshold values and when few rows are sampled in a yard. The error probability decreased more rapidly at the leaf level than at the sampling unit level, particularly when decisions were based on sampling one or two rows. The error probability can be kept below 5% at the leaflet and sampling unit scales when threshold values are 0.05 or greater and a minimum of three rows were sampled, and below 10% when threshold values are 0.1 or greater and a minimum of two rows were sampled. Interpreting Figure 5 must be done carefully because as the number of rows sampled in a yard increases so does the chance that a yard will be “fully” sampled. Since the “true” estimate of disease incidence in a yard is the mean incidence of the rows sampled in the yard, an incorrect decision is technically impossible once the yard is fully sampled under the current sampling plan. The proportion of yards that were fully sampled using the sampling plan described previously were 0.015, 0.18, 0.52, 0.75, and 0.83 for one, two, three, four, and five rows sampled, respectively.

## DISCUSSION

Incidence and heterogeneity of hop powdery mildew were characterized using data collected from single-row assessments and multiple-row or yard-level assessments; row-level analyses represented smaller scale assessments of the spatial pattern than yard-level assessments. Distributional and binary power law analyses, used to characterize pattern at the scale of the sampling unit, showed that disease incidence and heterogeneity of disease incidence were characteristic of a near-random disease pattern among sampling units collected within a single row. The spatial

pattern was found to be slightly more aggregated when analyses were conducted using information gathered from several rows within a single yard (i.e., yard-level analyses). Measures of spatial correlation, used to characterize larger scale patterns (i.e., plant-to-plant relationships within rows), indicated a low degree of spatial correlation. In instances where significant spatial correlation existed, it was typically confined to first-order correlations and there was no apparent relationship between the magnitude of the correlation and incidence or heterogeneity.

From a biological perspective, the data presents little evidence to suggest that epidemics begin from a few disease foci within yards and, in fact, argues that *P. macularis* perenniates successfully within yards at a significant proportion, despite current management practices. This is supported by ANOVA, which showed that the largest source of variation in disease incidence was among yards while the smallest source was among rows, indicating that incidence among yards was quite variable but within yards was relatively (but not entirely) homogenous. A similar situation exists in apple. Powdery mildew (*P. leucotricha*) epidemics start from conidia produced on infected primary shoots within the orchard (11) and lead to a random distribution of secondary powdery mildew on leaves (39). Random spatial patterns of incidence may be an attribute common to powdery mildews in perennial systems, particularly at the scale of the management system (i.e., individual yard, field, or orchard).

There is evidence of aggregation on a regional level. Regional effects were identified as a significant source of variation in both the covariance analysis and the nested ANOVA. Conceptually, the two analyses differ in that the covariance analysis is quantifying a factor's contribution to the extra-binomial variation of disease incidence, whereas ANOVA is quantifying a factor's contribution to the overall variation in incidence, i.e., binomial plus extra-binomial. Regional differences, however, were expected since hops is grown in two distinctly different climates in the Pacific

TABLE 6. Deviances and degrees of freedom (df) for five cases describing the relationship between incidence of hop powdery mildew at the leaf scale,  $p$ , and at the sampling unit scale,  $p_{su}$ , at either the row- or yard-level assessments in hop yards located in Washington and Oregon in 1999, 2000, and 2001

| Case <sup>a</sup> | Offset <sup>b</sup>  | $X^c$                       | Intercept/slope <sup>d</sup> | Row               |     | Yard  |     |
|-------------------|--|-----------------------------|------------------------------|-------------------|-----|-------|-----|
|                   |  |                             |                              | Dev. <sup>e</sup> | df  | Dev.  | df  |
| i                 | $\ln(n) + \text{CLL}(\hat{p})^f$   | —                           | —/—                          | 873.3             | 578 | 883.5 | 198 |
| ii                | $\ln \left[ -\ln \prod_{j=0}^{n-1} \left( \frac{1 - \hat{p} + j\theta}{1 + j\theta} \right) \right]$ | —                           | —/—                          | 467.5             | 578 | 233.0 | 198 |
| iii               | $\ln(n_0) + \text{CLL}(\hat{p})$   | —                           | —/—                          | 447.8             | 578 | 359.0 | 198 |
| iv                | $\ln(n_v) + \text{CLL}(\hat{p})$   | —                           | —/—                          | 398.0             | 578 | 250.4 | 198 |
| v                 | $\text{CLL}(\hat{p})$  | $\ln[\hat{p}(1 - \hat{p})]$ | $-\ln(a)/(1 - b)$            | 368.9             | 576 | 227.6 | 196 |

<sup>a</sup> Roman numeral refers to the “case” as described in the Appendix.

<sup>b</sup> Refers to terms in the model that are considered known.

<sup>c</sup> Independent variable; “—” indicates no independent variable in the model. All terms are offset, therefore no parameters are estimated.

<sup>d</sup> Theoretical value of the intercept and slope estimate; “—” indicates no intercept or slope estimated.

<sup>e</sup> Calculated deviance. For binomial data, values of the deviance equal to or less than the model degrees of freedom indicate a good fit. When comparing two models, if the reduction in deviance is greater than  $\chi^2_v$ , where  $v = df_1 - df_2$  is the absolute difference of the degrees of freedom for competing models, then the model with the smaller deviance is the better-fitting model.

<sup>f</sup>  $\text{CLL}(z) = \ln[-\ln(1 - z)]$ ; complementary log-log function.

TABLE 7. Analysis of variance table and variance components for the incidence of hop powdery mildew in hop yards located in Washington and Oregon in 2000

| Source of variation <sup>a</sup> | df     | Mean square <sup>b</sup> | $F$   | $P$   | Variance component <sup>c</sup> | % Total |
|----------------------------------|--------|--------------------------|-------|-------|---------------------------------|---------|
| Region                           | 4      | 48.540                   | 10.45 | 0.000 | 0.0069                          | 18.80   |
| Yard                             | 123    | 4.6466                   | 24.97 | 0.000 | 0.0159                          | 43.32   |
| Row                              | 378    | 0.1861                   | 16.32 | 0.000 | 0.0025                          | 6.81    |
| Sampling unit                    | 35,375 | 0.0114                   | ...   | ...   | 0.0114                          | 31.06   |
| Total                            | 35,800 | ...                      | ...   | ...   | ...                             | ...     |

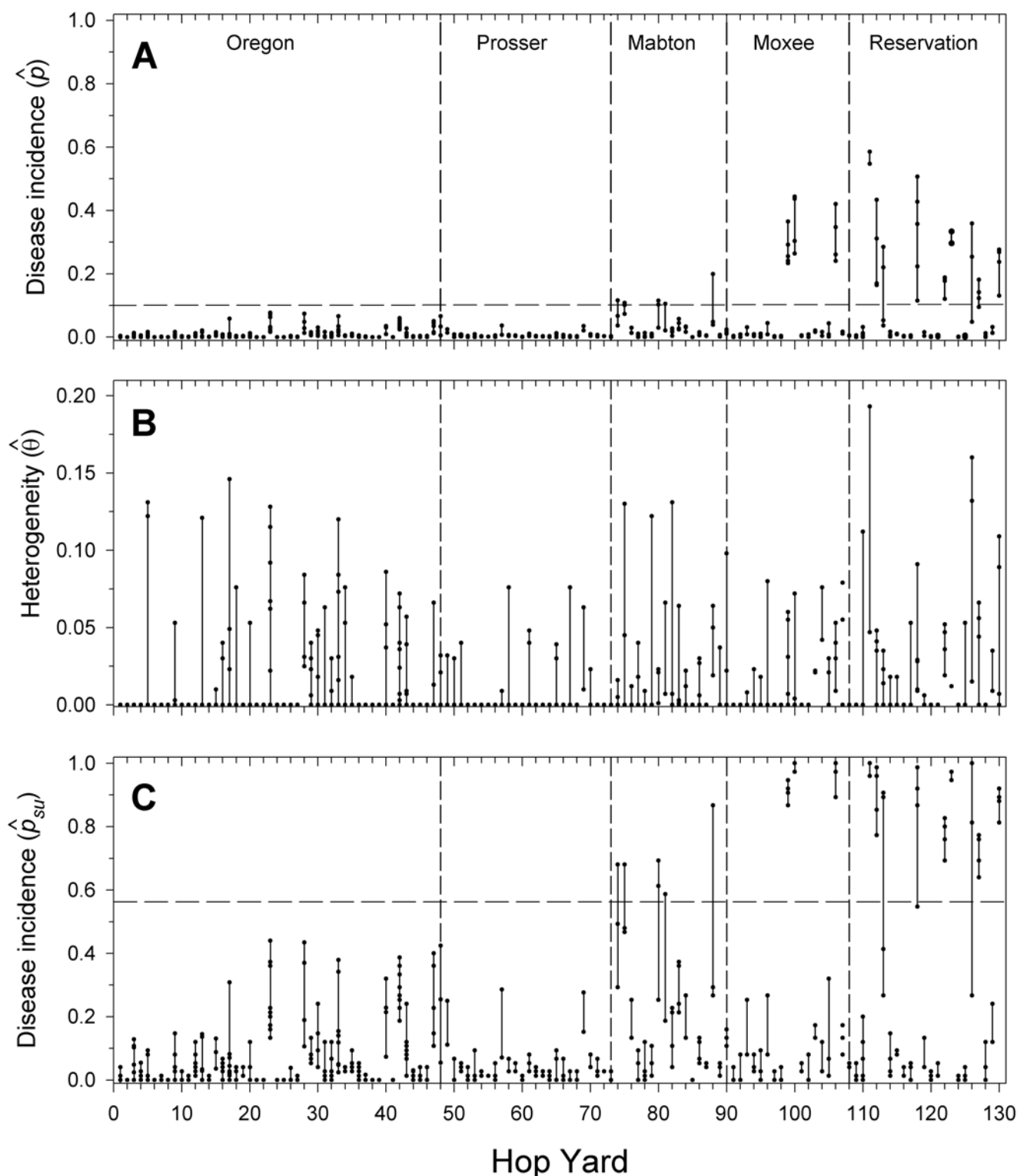
<sup>a</sup> Sampling units are nested within rows, rows are nested within yards, and yards are nested within regions. Five regions were considered: Oregon, Moxee, Reservation, Mabton, and Prosser. The latter four are located in the state of Washington.

<sup>b</sup> The mean squares for each source of variation are estimates of the following terms:  $\text{MS}_{\text{SU}} = s^2$ ;  $\text{MS}_{\text{row}} = s^2 + 70.52 s_{\text{row}}^2$ ;  $\text{MS}_{\text{yard}} = s^2 + 71.99 s_{\text{row}}^2 + 280.25 s_{\text{yard}}^2$ ; and  $\text{MS}_{\text{region}} = s^2 + 74.07 s_{\text{row}}^2 + 270.32 s_{\text{yard}}^2 + 6243.57 s_{\text{region}}^2$ .

<sup>c</sup> The variance components are obtained by substituting and rearranging terms accordingly.

Northwest. In Washington and southern Idaho, hops are grown in an arid climate, with March to June and September generally highly conducive to powdery mildew development. In Oregon and northern Idaho, the climate is much cooler and receives considerably more rainfall than Washington and southern Idaho, especially in the spring, with late June to September generally highly conducive to powdery mildew development. Differences

within Washington are less likely to be explained by climatic variation and more by the spatial proximity and arrangement of yards within a region and a region's location relative to the Reservation (considered the origin of introduction). Alternatively, one could speculate that powdery mildew epidemics begin in the western-most yards of the Reservation and Moxee (where powdery mildew intensity is generally highest) and that westward



**Fig. 4.** Range of hop powdery mildew incidence and heterogeneity values for single rows collected in the same yard on a single assessment date. **A**, Incidence and **B**, heterogeneity were assessed at the leaf level. **C**, Incidence was assessed at the sampling unit level. The broken line represents an artificial threshold incidence of 0.10 at the leaf level (corresponding to 0.57 at the sampling unit level). Analyses were conducted with 130 data sets collected in hop yards in Washington and Oregon in 2000.

winds carry conidia toward the east, depositing them randomly across yards. This could explain why flag shoots (the presumed source of initial inoculum) are difficult to find in yards outside of the Reservation that later developed high levels of powdery mildew. However, as *P. macularis* becomes established in a greater number of yards in the Yakima Valley, this explanation will be less likely and any subsequent differences may be attributed to climatic variations (25).

The yard-level analysis was done essentially to answer the question “Is it necessary to sample multiple rows in a yard or will single-row estimates of incidence provide the data necessary to make management decisions?” Combining the row estimates led to a distribution of incidence values that was more heterogeneous than the individual row estimates. From a mathematical perspective, this is primarily the result of yard-scale data sets having a larger proportion of sampling units with 0 or few diseased leaves than row-scale data sets, resulting in a positively skewed distribution that is not well described by the binomial distribution. Yard-level estimates of incidences “average in” high incidence values and encompass the variability inherent in most yards. Consequently, sampling formulae used to determine the number of sam-

pling units to observe in a yard should use parameters estimated from yard-level analyses; these estimates best characterize the variability at the management unit of concern. Once the number of sampling units to observe has been calculated, sampling should proceed over multiple rows in a single yard. Albeit the smallest component of variation, the variation among rows in a yard should not be ignored. Minimally, the number of sampling units observed should be divided among the three rows.

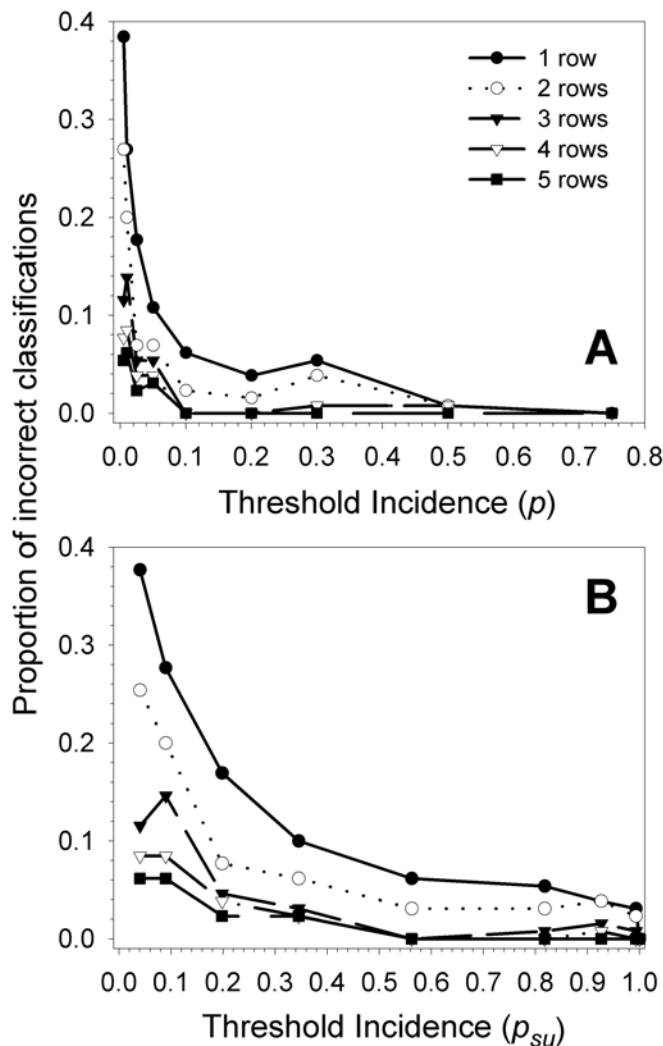
Moreover, our analyses show that greater emphasis should be placed on sampling a larger proportion of yards on any single farm; indirectly asserting that the time and effort committed to sampling in any one yard should be minimized. Sampling could be expedited if the disease status of the sampling unit could be assessed more rapidly. This was addressed with the hierarchical analysis characterizing the relationship between incidences at the leaf and sampling unit scales (37). The good fit to the model(s) suggests that characterizing the disease status of a sampling unit could replace enumerating the disease status of the 10 individual leaves as a practical way of sampling. In so doing, the sampling approach is similar. The sampler approaches each plant with the intent of sampling 10 leaves, but differs in that once the first diseased leaf is discovered, the sampler moves on to the next sampling unit of 10 leaves. This approach is intended to mimic whole-plant sampling and, indeed, assessing the incidence of a plant rather than attempting to single out 10 leaves may perform similarly (but was not tested). To have developed the scheme using entire plants would have required us to sample every leaf on a hop plant (5.5 m tall by 2 m wide), which is unrealistic.

Characterizing the disease status of a sampling unit rather than the individual leaves comes with the cost of a loss of accuracy. If one were to interchange the axes in Figure 3, allowing  $p_{su}$  to serve as the independent variable and  $p$  as the dependent variable, it is clear that for any value of  $p_{su}$ , particularly at low values of incidence, a relatively large range of  $p$  values are possible. The practical question is “What sampling procedure minimizes the risk of making a management error?” That is, one must weigh sampling a greater proportion of yards and obtaining estimates of incidence that are less accurate than what could be obtained by sampling fewer yards more intensively. Given that a large proportion of the variability can be attributed to yard-to-yard variation, and knowing that growers are managing multiple yards at any given time, sampling quickly in any single yard so that the total number of yards sampled on the farm can be increased is the most desirable approach.

In conclusion, growers and crop specialists often decide based on a limited sampling whether or not an action, such as a fungicide application, is necessary. Hop yards are generally large. Thus, it is difficult to know without any prior knowledge of the spatial structure of disease incidence, how many samples should be taken, at what scale measurements of disease should be made, and what proportion of the yard samples should be taken so incidence can be estimated with reasonable certainty. The foundation for answering these questions was developed in this research. The question of how many samples or, practically, what is the *minimum* sample size needed to make an informed decision can be answered based on the information collected here with the development and validation of a sampling plan, but this is beyond the scope of this paper and is the focus of future research. Moreover, what is a reasonable incidence threshold to build a sampling plan around has yet to be addressed; this is perhaps one of the most important components in the design of efficient sampling procedures (27) and will be addressed in future research.

## APPENDIX

Farrington (8) showed that the binomial model (equation 3) can be reformulated in a generalized linear model to permit direct curve fitting (1), allowing for explicit consideration of the bino-



**Fig. 5.** Proportion of hop yards classified incorrectly as being under a powdery mildew incidence threshold of **A**,  $p = 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.5$ , or  $0.75$  (leaf level) and **B**,  $p_{su} = 0.0406, 0.0901, 0.1982, 0.3455, 0.5624, 0.8177, 0.9265, 0.9933$ , and  $1.0$  (sampling unit level) as a function of the number of rows sampled. The value of  $p_{su}$  was calculated using equation A2 with estimated slope of  $1 - b = -0.0816$  and intercept  $-\ln(a) = 1.863$  as described in the text. Analyses were conducted with 130 data sets collected in hop yards in Washington and Oregon in 2000.

mial error structure of  $\hat{p}_{su}$ . If  $CLL(z) = \ln[-\ln(1 - z)]$  is the complementary log-log transformation of  $z$ , then equation 3 can be rearranged such that

$$CLL(\tilde{p}_{su}) = \ln(n_{\bullet}) + CLL(\hat{p}) \quad (A1)$$

This is an equation for a straight line with a slope of 1 and intercept of  $\ln(n_{\bullet})$ , where  $n_{\bullet}$  is replaced by either  $n$ ,  $n_D$ ,  $n_{\delta}$ , or  $n_v$ , as defined previously. As Turechek and Madden (37) indicate, equation A1 can be viewed as a simple reformulation of equation 3, thus requiring no curve fitting to estimate the parameters (because they are already known), or it could be viewed as a statistical model for  $CLL(\tilde{p}_{su})$ : $CLL(\hat{p})$ ; two of these models were considered. The first was to assume a known slope of 1 and an unknown intercept equivalent to  $\ln(n_{\bullet})$ . In this case,  $\ln(n_{\bullet})$  was not necessarily an effective sample size, but a parameter describing the  $CLL(\tilde{p}_{su})$ : $CLL(\hat{p})$  relationship. For the second model,  $n_{\delta}$  (equation 6) was substituted for  $n_{\bullet}$  into equation A1 and, after some rearrangement,

$$CLL(\tilde{p}_{su}) = -\ln(a) + (1 - b) \cdot \ln[\hat{p}(1 - \hat{p})] + CLL(\hat{p}) \quad (A2)$$

To fit this model,  $CLL(\hat{p})$  was subtracted from both sides of equation A2, i.e.,  $CLL(\hat{p})$  is an "offset" variable,  $\ln[\hat{p}(1 - \hat{p})]$  was treated as an independent variable with slope  $(1 - b)$  and intercept  $-\ln(a)$ . When the parameters  $a$  and  $b$  were estimated through direct curve fitting, their interpretation is similar to, but not exactly as in equation 1.

The deviance, analogous to the residual sum of squares for normal-theory regression, was calculated for five candidate models; four required no direct curve fitting or parameter estimation, one model did. The four cases requiring no parameter estimations were (i) binomial (random) distribution of diseased leaves (equation A1 with  $n_{\bullet} = n$ ); (ii) beta-binomial model (equation 4 formulated as a CLL [see equation 15 in reference 37]); (iii) effective sample size assuming  $D$  varies systematically with  $p$  according to the binary power law (equation A1 with  $n_{\bullet} = n_{\delta}$ ); and (iv) effective sample size of Madden and Hughes (21) with variable heterogeneity (equation A1 with  $n_{\bullet} = n_v$ ). The model requiring curve fitting, case v, was equation A2 fit as described above.

The statistical package GLIM was used to estimate the deviance. For cases i to v, where no parameters were estimated, the GLIM method of Aitkin et al. (1) (page 177 of literature citation 1) was used to determine the deviances. The five models were first fit to the individual row data for the 3 years combined and then to the yard data for the 2 years combined.

## ACKNOWLEDGMENTS

Salaries and research support were provided by federal and state funds apportioned to the New York State Agriculture Experiment Station, USDA-ARS (CRIS 303-5358-22000-024-00D), and Oregon State University, and grants from the Hop Research Council, Washington Commission of Pesticide Registration, and Washington Hop Commission. We thank C. Bradimer, I. Cantrell, J. DiLeone, K. Donahue, N. Gargus, J. Heffner, S. McMullen, M. Meier, S. Scheuerell, M. Scott, T. Sechler, H. Sharpless, C. Soskis, A. Taylor, and R. Whitmore for help with disease assessment.

## LITERATURE CITED

- Aitkin, M., Anderson, D., Francis, B., and Hinde, J. 1989. Statistical Modeling in GLIM. Clarendon Press, Oxford.
- Binns, M. R., Nyrop, J. P., van der Werf, W. 2000. Sampling and Monitoring in Crop Protection: The Theoretical Basis for Developing Practical Decision Guides. CABI Publishing, Oxon, UK.
- Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.
- Cliff, A. D., and Ord, J. K. 1981. Spatial Processes: Models and Applications. Pion Ltd., London.
- Dale, M. R. T. 1999. Spatial Pattern Analysis in Plant Ecology. Cambridge University Press, Cambridge, UK.
- Dallot, S., Gottwald, T., Labonne, G., and Quiot, J.-B. 2003. Spatial pattern analysis of Sharka disease (*Plum pox virus* strain M) in peach orchards of southern France. *Phytopathology* 93:1543-1552.
- Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. *Ecology* 74:1646-1658.
- Farrington, C. P. 1992. Estimating prevalence by group testing using generalized linear models. *Stat. Med.* 11:1591-1597.
- Ferrandino, F. 1989. A distribution-free method for estimating the effects of aggregated plant damage on crop yield. *Phytopathology* 79:1229-1232.
- Gubler, W. D., Rademacher, M. R., Vasquez, S. J., and Thomas, C. S. 1999. Control of powdery mildew using the UC Davis powdery mildew risk index. *APSnets* Feature January 1999. Published online by The American Phytopathological Society, St. Paul, MN.
- Hickey, K. D., and Yoder, K. S. 1990. Powdery mildew. Pages 9-10 in: *Compendium of Apple and Pear Diseases*. A. L. Jones and H. S. Aldwinckle, eds. The American Phytopathological Society, St. Paul, MN.
- Hughes, G. 1988. Spatial heterogeneity in crop loss assessment models. *Phytopathology* 78:883-884.
- Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. *Plant Pathol.* 41:657-660.
- Hughes, G., and Madden, L. V. 1995. Some methods allowing for aggregated patterns of disease incidence in the analysis of designed experiments. *Plant Pathol.* 44:927-943.
- Hughes, G., Madden, L. V., and Munkvold, G. P. 1996. Cluster sampling for disease incidence data. *Phytopathology* 86:132-137.
- Hughes, G., McRoberts, N., Madden, L. V., and Gottwald, T. R. 1997. Relationship between disease incidence at two levels in a spatial hierarchy. *Phytopathology* 87:542-550.
- Kish, L. 1965. Survey Sampling. John Wiley & Sons, New York.
- Ludwig, J. A., and Reynolds, J. F. 1988. Statistical Ecology: A Primer on Methods and Computing. John Wiley & Sons, New York.
- Madden, L. V., and Hughes, G. 1994. BBD-computer software for fitting the beta-binomial distribution to disease incidence data. *Plant Dis.* 78:536-540.
- Madden, L. V., and Hughes, G. 1995. Plant disease incidence: Distributions, heterogeneity, and temporal analysis. *Annu. Rev. Phytopathol.* 33:529-564.
- Madden, L. V., and Hughes, G. 1999. An effective sample size for predicting plant disease incidence in a spatial hierarchy. *Phytopathology* 89:770-781.
- Madden, L. V., and Nutter, F. W., Jr. 1995. Modeling crop losses at the field scale. *Can. J. Plant Pathol.* 17:124-137.
- Madden, L. V., Turechek, W. W., and Nita, M. 2002. Evaluation of generalized linear mixed models for analyzing disease incidence data obtained from designed experiments. *Plant Dis.* 86:316-325.
- Mahaffee, W. F., Thomas, C. S., Turechek, W. W., Ocamb, C. M., Nelson, M., and Gubler, W. D. 2003. Responding to an introduced pathogen: *Podosphaera macularis* (hop powdery mildew) in the Pacific Northwest. Online. *Plant Health Progress* doi:10.1094/PHP-2003-1113-07-RV.
- Mahaffee, W. F., Turechek, W. W., and Ocamb, C. M. 2003. Effect of variable temperature on infection severity of *Podosphaera macularis* on hops. *Phytopathology* 93:1587-1592.
- Neve, R. A. 1991. Hops. Chapman and Hall, London.
- Nyrop, J. P., Binns, M. R., and van der Werf, W. 1999. Sampling for IPM decision making: Where should we invest our time and resources? *Phytopathology* 89:1104-1111.
- Ocamb, C., Klein, R., Barbour, J., Griesbach, J., and Mahaffee, W. 1999. First report of hop powdery mildew in the Pacific Northwest. *Plant Dis.* 83:1072.
- Pethybridge, S. J., and Turechek, W. W. 2003. Analysis of the association among three viruses infecting hop in Australia. *Plant Pathol.* 52:158-167.
- Rao, J. N. K., and Scott, A. J. 1992. A simple method for the analysis of clustered binary data. *Biometrics* 48:577-585.
- Royle, D. J. 1978. Powdery mildew of the hop. Pages 381-409 in: *The Powdery Mildews*. D. M. Spencer, ed. Academic Press, London.
- Scott, J. B., Hay, F. S., Wilson, C. R., Cotterill, P. J., and Fitt, A. J. 2003. Spatiotemporal analysis of epiphytotics of downy mildew of oilseed poppy in Tasmania, Australia. *Phytopathology* 93:752-757.
- Tomlan, M. A. 1992. Tinged with Gold: Hop Culture in the United States. University of Georgia Press, Athens.
- Turechek, W. W., Ellis, M. A., and Madden, L. V. 2000. Sequential sampling for incidence of Phomopsis leaf blight of strawberry. *Phytopathology* 91:336-347.
- Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis of

- strawberry leaf blight in perennial production systems. *Phytopathology* 89:421-433.
36. Turecek, W. W., and Madden, L. V. 1999. Spatial pattern analysis and sequential sampling for the incidence of leaf spot on strawberry in Ohio. *Plant Dis.* 83:992-1000.
  37. Turecek, W. W., and Madden, L. V. 2003. A generalized linear modeling approach for characterizing disease incidence in a spatial hierarchy. *Phytopathology* 93:458-466.
  38. Turecek, W. W., Mahaffee, W. F., and Ocamb, C. M. 2001. Development of management strategies for hop powdery mildew in the Pacific Northwest. Online. *Plant Health Progress* doi:10.1094/PHP-2001-0313-01-RS.
  39. Xu, X.-M., and Madden, L. V. 2002. Incidence and density relationships of powdery mildew on apple. *Phytopathology* 92:1005-1014.
  40. Young, L. J., and Young, J. H. 1998. *Statistical Ecology*. Kluwer Academic Publishers, Boston.